N. Lange,[1,2] P. A. Bandettini,[2] J. R. Baker,[2] B. R. Rosen[2]
[1]National Institutes of Health, Bethesda, MD [2]MGH-NMR Center, Charlestown, MA

INTRODUCTION Many activation detection methods have been proposed recently for human brain mapping by functional MRI. Six such methods are compared here using actual data. While limited in their generality, actual comparative data are preferable to computer simulations: To simulate, one must presume to know the stochastic image generation mechanism completely, which is unrealistic, and, moreover, existing statistical theory under many such assumptions makes actual simulation unnecessary.

METHODS Checkerboards reversing at 8 Hz were presented alternately to right and left visual hemifields. The stimulus avoided the central foveal region in order to achieve maximal separation of right and left brain hemispheres. Two oblique slices parallel to the calcarine fissure were imaged, one superior, one inferior. All data were obtained using an asymmetric spin echo sequence (ASE; TR = 1000 ms, TE = 70 ms, $\Delta\tau = -25$ ms; see [1]). Voxel size was $3.125 \times 3.125 \times 6$ mm. Sixty-four images were collected under right hemifield (left hemisphere) stimulation, 64 left hemifield (right hemisphere), and so forth for a total of 512 images. This design was repeated at 2 Hz and 4 Hz flashing frequency during the same session. The spatial time series were detrended by removal of a temporal least-squares regression line. Resulting temporal residuals were smoothed by an adaptive bandwidth kernel. There was no spatial smoothing. On-screen animation revealed no appreciable head motion artifact. *Prior to any statistical analysis*, two of us (NL, PAB) independently declared as "truly activated" those voxels whose time courses had features that we, as "experts", associated with activation. In this manner, 73 voxels were labelled as "truly activated", and 150 neighboring voxels labelled as "truly not activated" (Figure 1). These 223 voxels constituted an operating "gold standard" by which to compare six activation detection methods by their receiver operating characteristic curves.

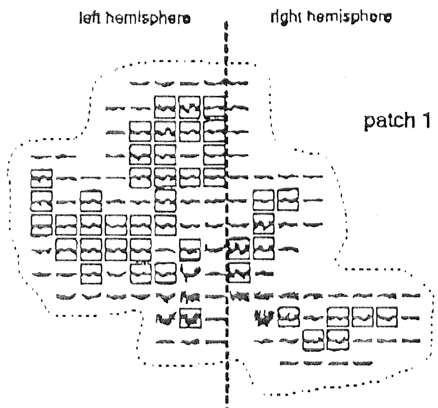

Figure 1. One of two sets of spatial time series used; note negative correlation between different hemispheres.

The six methods, each applied voxel-by-voxel, are: (1) a two-sample t-test, (2) a Gram-Schmidt correlation analysis [2], (3) a paired t-test, (4) a Kolmogorov-Smirnov test, (5) a normalized cross-covariance method [3], and (6) a modified version of (5), see [4]. A two-sample t-test is equivalent to Gram-Schmidt, which in turn is equivalent to linear regression on a binary covariate when using a canonical (orthonormal) design matrix. A paired t-test is a simple version of repeated measures anova with a binary factor. In (5), temporal smoothness is estimated as $\varsigma_t^2 = \sigma_t^2/(2\sigma_{dt}^2)$, whereas in (6) this is estimated as $-0.25/\ln\left[1-\sigma_{dt}^2/(2\sigma_t^2)\right]$. Both (5) and (6) were fit in the frequency domain under a stationary and isotropic Gaussian autocovariance model.

RESULTS Receiver operating characteristic curves for each method applied to the first 256 images are shown in Figure 2. Since right hemisphere is negatively correlated with left hemisphere, by design, absolute values of test statistics were used throughout. Results for methods (5) and (6) were virtually identical at this TR, hence only (5) is shown. All methods have an approximately 60% true activation detection ratio (statistical power) at about 5% false activation detection ratio (type I error). Overall, the normalized cross-covariance method appears to dominate the K-S test. K-S appears to dominate the t-tests, due to heavy tails, unequal variances and serial dependence. In fact, the estimated (non-integral) degrees of freedom were as low as 192.3, and serial correlation of roughly 0.96 drops to 0.50 after about lag 6. Comparable curves for spatial time series twice as long (512 images) showed poorer performance, as did the 2 Hz and 4 Hz designs, due in part to increased nonstationary effects (eg. blinking).

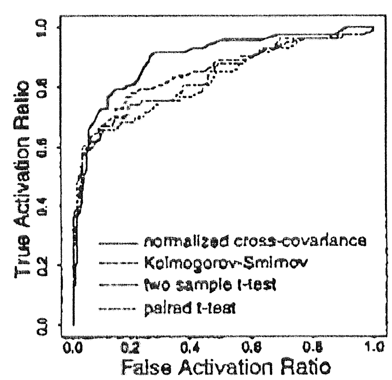

Figure 2.

CONCLUSIONS and DISCUSSION These data suggest that an isotropic Gaussian autocovariance model [3] may yield better overall performance than a model-free approach, such as K-S. However, while this aspect of the model appears valid [5], the assumption of constant hemodynamic delay in [3] does not. The normalized cross-covariance approach can be improved by allowing hemodynamic responses to vary spatially [5].

REFERENCES

1. Baker, J. R., Stern, C. E., Hoppel, B. E., et al. Technical report, MGH-NMR Center, 1995.
2. Bandettini, P. A., Jesmanowicz, A., Wong, E. C., Hyde, J. S., *Magn. Reson. Med.*, 30, 161–173, 1993.
3. Friston, K. J., Jezzard, P., Turner, R. *Human Brain Mapping*, 1, 153–174, 1994.
4. Forman, S. D., Cohen, J. D., Fitzgerald, M. et al. In press, *Magn. Reson. Med.*, 1995.
5. Lange, N., Zeger, S. L. Technical report, NIH/NINDS, 1995.